ATTACHMENT 17

Source and Accuracy of Estimates for the June 1998 CPS
Microdata File on Fertility and Birth Expectations

SOURCE OF DATA

The data in this microdata file come from the June 1998 Current Population Survey (CPS). The Bureau of the Census conducts the survey every month, although this file has only June data. The June survey uses two sets of questions, the basic CPS and the supplement.

Basic CPS. The basic CPS collects primarily labor force data about the civilian noninstitutional population. Interviewers ask questions concerning labor force participation about each member 15 years old and over in every sample household.

Sample design. The present CPS sample was selected from the 1990 Decennial Census files with coverage in all 50 states and the District of Columbia. The sample is continually updated to account for new residential construction. The United States was divided into 2,007 geographic areas. In most states, a geographic area consisted of a county or several contiguous counties. In some areas of New England and Hawaii, minor civil divisions are used instead of counties. A total of 754 geographic areas were selected for sample. About 50,000 occupied households are eligible for interview every month. Interviewers are unable to obtain interviews at about 3,200 of these units. This occurs when the occupants are not found at home after repeated calls or are unavailable for some other reason.

Since the introduction of the CPS, the Bureau of the Census has redesigned the CPS sample several times. These redesigns have improved the quality and accuracy of the data and have satisfied changing data needs. The most recent changes were completely implemented in July 1995.

June Supplement. In addition to the basic CPS questions, interviewers asked supplementary questions in June about fertility of women between 15 and 44 years of age.

Estimation procedure This survey's estimation procedure adjusts weighted sample results to agree with independent estimates of the civilian noninstitutional population of the United States by age, sex, race, Hispanic/non-Hispanic origin, and state of residence. The adjusted estimate is called the post-stratification ratio estimate. The independent estimates are calculated based on information from four primary sources:

- The 1990 Decennial Census of Population and Housing.
- An adjustment for undercoverage in the 1990 census.
- Statistics on births, deaths, immigration, and emigration.
- Statistics on the size of the Armed Forces.

The independent population estimates include some, but not all, undocumented immigrants.

ACCURACY OF THE ESTIMATES

Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. A sample survey estimate has two possible types of errors: sampling and nonsampling. The accuracy of an estimate depends on both types of errors, but the full extent of the nonsampling error is unknown. Consequently, one should be particularly careful when interpreting results based on a relatively small number of cases or on small differences between estimates. The standard errors for CPS estimates primarily indicate the magnitude of sampling error. They also partially measure the effect of some nonsampling errors in responses and enumeration, but do not measure systematic biases in the data. (Bias is the average over all possible samples of the differences between the sample estimates and the desired value.)

Nonsampling Variability. There are several sources of nonsampling errors including the following:

Inability to obtain information about all cases in the sample.
Definitional difficulties.
Differences in the interpretation of questions.
Respondents' inability or unwillingness to provide correct information.
Respondents' inability to recall information.
Errors made in data collection such as in recording or coding the data.
Errors made in processing the data.
Errors made in estimating values for missing data.
Failure to represent all units with the sample (undercoverage).

For the June 1998 basic CPS, the nonresponse rate was 6.7% and for the supplement the nonresponse rate was an additional 4.1% for a total supplement nonresponse rate of 10.5%.

CPS undercoverage results from missed housing units and missed persons within sample households. Overall CPS undercoverage is estimated to be about 8 percent. CPS undercoverage varies with age, sex, and race. Generally, undercoverage is larger for males than for females and larger for Blacks and other races combined than for Whites. As described previously, ratio estimation to independent age-sex-race-Hispanic population controls partially corrects for the bias due to undercoverage. However, biases exist in the estimates to the extent that missed persons in missed households or missed persons in interviewed households have different characteristics from those of interviewed persons in the same age-sex-race-origin-state group.

A common measure of survey coverage is the coverage ratio, the estimated population before poststratification divided by the independent population control. Table A shows CPS coverage ratios for age-sex-race groups for a typical month. The CPS coverage ratios can exhibit some variability from month to month. Other Census Bureau household surveys experience similar coverage.

Table 1. CPS Coverage Ratios

	Non-Black		Black		All Persons		
Age	M	F	M	F	M	F	Total
0-14	0.929	0.964	0.850	0.838	0.916	0.943	0.929
15	0.933	0.895	0.763	0.824	0.905	0.883	0.895
16-19	0.881	0.891	0.711	0.802	0.855	0.877	0.866
20-29	0.847	0.897	0.660	0.811	0.823	0.884	0.854
30-39	0.904	0.931	0.680	0.845	0.877	0.920	0.899
40-49	0.928	0.966	0.816	0.911	0.917	0.959	0.938
50-59	0.953	0.974	0.896	0.927	0.948	0.969	0.959
60-64	0.961	0.941	0.954	0.953	0.960	0.942	0.950
65-69	0.919	0.972	0.982	0.984	0.924	0.973	0.951
70+	0.993	1.004	0.996	0.979	0.993	1.002	0.998
15+	0.914	0.945	0.767	0.874	0.898	0.927	0.918
0+	0.918	0.949	0.793	0.864	0.902	0.931	0.921

For additional information on nonsampling error including the possible impact on CPS data when known, refer to Statistical Policy Working Paper 3, An Error Profile: Employment as Measured by the Current Population Survey, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, 1978 and Technical Paper 40, The Current Population Survey: Design and Methodology, Bureau of the Census, U.S. Department of Commerce.

Comparability of data Data obtained from the CPS and other sources are not entirely comparable. This results from differences in interviewer training and experience and in differing survey processes. This is an example of nonsampling variability not reflected in the standard errors. Use caution when comparing results from different sources.

A number of changes were made in data collection and estimation procedures beginning with the January 1994 CPS. The major change was the use of a new questionnaire. The questionnaire was redesigned to measure the official labor force concepts more precisely, to expand the amount of data available, to implement several definitional changes, and to adapt to a computer-assisted interviewing environment. The March supplemental income questions were also modified for adaptation to computer-assisted interviewing, although there were no changes in definitions and concepts. Due to these and other changes, one should use caution when comparing estimates from data collected in 1994 and later years with estimates from earlier years.

Caution should also be used when comparing data from this microdata file, which reflects 1990 census-based population controls, with microdata files from December 1993 and earlier years, which reflect 1980 census-based population controls. This change in population controls had relatively little impact on summary measures such as means, medians, and percentage distributions. It did have a significant impact on levels. For example, use of 1990 based population controls results in about a 1-percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 1994 and later years will differ from those for earlier years by ore than what could be attributed to actual changes in the population. These differences could be disproportionately greater for certain subpopulation groups than for the total population.

Since no independent population control totals for persons of Hispanic origin were used before 1985, compare Hispanic estimates over time cautiously.

Caution should also be exercised when using metropolitan/nonmetropolitan estimates during the redesigned CPS sample phase-in period from April 1994 through June 1995. During this phase-in period, CPS data were collected from sample designs based on both the 1980 and 1990 censuses. While most CPS estimates have been unaffected by this mixed sample, metropolitan/nonmetropolitan estimates have been affected. Sample cases from the new design were recoded to reflect the 1980 metropolitan/nonmetropolitan definitions to allow the estimates to be comparable with earlier data. The gross error rate for the conversions of central cities/suburbs is not expected to exceed 5%.

Note When Using Small Estimates Because of the large standard errors involved, summary measures (such as medians and percentage distributions) would probably not reveal useful information when computed on a smaller base than 75,000.

Take care in the interpretation of small differences. For instance, even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Sampling Variability. Sampling variability is variation that occurred by chance because a sample was surveyed rather than the entire population. Standard errors, as calculated by methods described later in **Standard Errors and Their Use**, are primarily measures of sampling variability, although they may include some nonsampling error.

Standard Errors and Their Use. A number of approximations are required to derive, at a moderate cost, standard errors applicable to all the estimates in this microdata file. Instead of providing an individual standard error for each estimate, parameters are provided to calculate standard errors for various types of characteristics. These parameters are listed in Tables 2-4.

The sample estimate and its standard error enable one to construct a confidence interval, a range that would include the average result of all possible samples with a known probability. For example, if all possible samples were surveyed under essentially the same general conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

A particular confidence interval may or may not contain the average estimate derived from all possible samples. However, one can say with specified confidence that the interval includes the average estimate calculated from all possible samples.

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. One common type of hypothesis is that the population parameters are different. An example of this would be comparing the percentage of employed males 20 to 24 years old working part time to the percentage of employed females in the same age group who were part-time workers. An illustration of this is included in the following pages.

Tests may be performed at various levels of significance. A significance level is the probability of concluding that the characteristics are different when, in fact, they are the same. To conclude that two parameters are different at the 0.10 level of significance the absolute value of the estimated difference between characteristics must be greater than or equal to 1.645 times the standard error of the difference.

The Census Bureau uses 90-percent confidence intervals and 0.10 levels of significance to determine statistical validity. Consult standard statistical textbooks for alternative criteria.

For information on calculating standard errors for labor force data from the CPS which involve quarterly or yearly averages, changes in consecutive quarterly or yearly averages, consecutive month-to-month changes in estimates, and consecutive year-to-year changes in monthly estimates, see "Explanatory Notes and Estimates of Error: Household Data" in the corresponding *Employment and Earnings* published by the Bureau of Labor Statistics.

Standard errors of estimated numbers The approximate standard error, s_x, of an estimated number from this microdata file can be obtained using this formula:

$$s_x = \sqrt{ax^2 + bx}$$

Formula (1)

Here x is the size of the estimate and a
and b are the parameters in Table 2 or 3 associated with the particular type of characteristic. When calculating standard errors from cross-tabulations involving different characteristics, use the set of parameters for the characteristic which will give the largest standard error.

Illustration

Suppose there were 5,360,000 unemployed females in the civilian labor force. Use the appropriate parameters from Table 2 and Formula (1) to get

Number, x	2,900,000
a parameter	-0.000018
b parameter	2,957
standard error	101,000
90% conf. int.	3,334,000 to 3,666,000

The standard error is calculated as

$$s_x = \sqrt{-0.000018 \times 2{,}900{,}000^2 + 2{,}957 \times 2{,}900{,}000} = 101{,}000$$

the 90-percent confidence interval is calculated as 3,334,000 ± 1.645 x 101,000.

A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

<u>Standard Errors of Estimated Percentages</u>. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends on the size of the percentage and its base. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. When the numerator and denominator of the percentage are in different categories, use the parameter from Table 2 or 3 indicated by the numerator.

The approximate standard error, $s_{x,p}$, of an estimated percentage can be obtained by use of the formula

$$s_{x,p} = \sqrt{\frac{b}{x}\, p(100 - p)}$$

Formula (2)

Here x is the total number of persons, families, households, or unrelated individuals in the base of the percentage, p is the percentage ($0 \le p \le 100$), and b is the parameter in Table 2 or 3 associated with the characteristic in the numerator of the percentage.

<u>Illustration</u>.

Suppose that 6.1 percent of the 60,526,000 women 15-44 years old had a child in the last year. Use the appropriate parameter from Table 3 and formula (2) to get

Percentage, p	6.1
Base, x	60,526,000
b parameter	2,241
Standard error	0.15
90% conf. int.	5.9 to 6.3

The standard error is calculated as

$$s_{x,p} = \sqrt{(2{,}241/60{,}526{,}000)(6.1)(100.0 - 6.1)} = 0.14$$

The 90-percent confidence interval for the percentage of women 15-44 years old who had a child in the last year is calculated as 6.14 ± 1.645x0.14.

Standard Error of a Difference. The standard error of the difference between two sample estimates is approximately equal to

$$s_{x-y} = \sqrt{s_x^2 + s_y^2}$$

Formula (3)

where s_x and s_y are the standard errors of the estimates, x and y. The estimates can be numbers, percentages, ratios, etc. This will result in accurate estimates of the standard error of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same area. However, if there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

Illustration

Suppose that of 6,285,000 employed men between 20-24 years of age, 1,516,000 or 24.1 percent were part-time workers, and of the 5,824,000 employed women between 20-24 years of age, 2,169,000 or 37.2 percent were part-time workers. Use the appropriate parameters from Table 2 and formulas (3) and (4) to get

	x	y	difference
Percentage, p	24.1	37.2	13.1
Number, x	6,285,000	5,824,000	-
b parameter	2,764	2,530	-
Standard error	0.9	1.0	1.3
90% conf. int.	22.6 to 25.6	35.6 to 38.8	11.0 to 15.2

The standard error of the difference is calculated as

$$s_{x-y} = \sqrt{0.9^2 + 1.0^2} = 1.3$$

The 90-percent confidence interval around the difference is calculated as 13.1 ± 1.645×1.3. Since this interval does not include zero, we can conclude with 90 percent confidence that the percentage of part-time women workers between 20-24 years of age is greater than the percentage of part-time men workers between 20-24 years of age.

Standard error of a fertility ratio. The standard error of a fertility ratio is a function of the number of children ever born per 1,000 women and the number of women in a given category. The formula for the standard error of a fertility ratio is

$$s_{x,y} = x\sqrt{a + \frac{b}{x\,y} + \frac{c}{1{,}000y}}$$

Formula (4)

where a, b and c are the parameters from Table 4, x is the number of children ever born or expected per 1,000 women and y is the number of women, in thousands. This formula should be used when calculating standard errors for estimates involving the possibility of more than one event per woman, i.e., number of children ever born. For data involving at most one event per woman, convert the ratio to a percentage and use formula (2) and the parameters in Table 2 or 3 to calculate the standard errors.

Illustration.

Suppose that 9,252,000 ever-married women 40-44 years old had 2,087 children ever born per 1,000 women. Use formula (4) and the parameters in Table 4 to get

Children Ever Born, x	2,087
Base, 1,000y	9,252,000
a parameter	+0.0000014
b parameter	901
c parameter	1,644
Standard error	31
90% conf. int.	2,036 to 2,138

The standard error is calculated as

$$s_{x,y} = 2{,}087\sqrt{0.0000014 + \frac{901}{2{,}087 \times 9{,}252} + \frac{1{,}644}{1{,}000 \times 9{,}252}} = 31$$

The 90-percent confidence interval is from 2,036 to 2,138 children ever born per 1,000 women (i.e., 2,087 ± 1.645×31). A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

Standard Error of a Ratio. Certain estimates may be calculated as the ratio of two numbers. The standard error of a ratio, x/y, may be computed using

$$s_{x/y} = \frac{x}{y}\sqrt{\left[\frac{s_x}{x}\right]^2 + \left[\frac{s_y}{y}\right]^2 - 2r\frac{s_x}{x}\frac{s_y}{y}}$$

Formula (5)

The standard error of the numerator, s_x, and that of the denominator, s_y, may be calculated using formulas described earlier. In formula (5), r represents the correlation between the numerator and the denominator of the estimate.

For one type of ratio, the denominator is a count of families or households and the numerator is a count of persons in those families or households with a certain characteristic. If there is at least one person with the characteristic in every family or household, use 0.7 as an estimate of r. An example of this type is the mean number of children per family with children.

For all other types of ratios, r is assumed to be zero. If r is actually positive (negative), then this procedure will provide an overestimate (underestimate) of the standard error of the ratio. Examples of this type are the mean number of children per family and the poverty rate.

NOTE: For estimates expressed as the ratio of x per 100 y or x per 1,000 y, multiply formula (5) by 100 or 1,000, respectively, to obtain the standard error.

Illustration.

Suppose there are 37,378,000 ever-married women 15-44 years old and 22,846,000 never-married women 15-44 years old. The ratio of ever-married women, x, to never-married women, y, is 1.64. Use the appropriate parameters from Table 3 and equations (1) and (5) to get

	x	y	ratio
Estimate	37,378,000	22,846,000	1.64
a parameter	-0.000025	-0.000025	-
b parameter	5,211	5,211	-
Standard error	400,000	326,000	0.03
90% conf. int.	36,200,000 to 38,036,000	22,310,000 to 23,382,000	1.59 to 1.69

Using formula (5) with r = 0, the estimate of the standard error is

$$s_{x/y} = \frac{37{,}378{,}000}{22{,}846{,}000}\sqrt{\left[\frac{400{,}000}{37{,}378{,}000}\right]^2 + \left[\frac{326{,}000}{22{,}846{,}000}\right]^2} = 0.03$$

The 90-percent confidence interval is calculated as 1.64 ± 1.645x0.03.

Standard errors for region, state and nonmetropolitan estimates. Multiply the parameters in Tables 2, 3 and 4 by the factors in Tables 5 and 6 to get state, region and nonmetropolitan parameters for labor force and fertility estimates.

Table 2. Parameters for Computation of Standard Errors for Labor Force Characteristics: June 1998		
Characteristic	a	b
Labor Force and Not In Labor Force Data Other than Agricultural Employment and Unemployment		
Total [1]		
Men [1]	-0.000018	2,985
Women	-0.000033	2,764
Both sexes, 16 to 19 years	-0.000030	2,530
	-0.000172	2,545
White [1]	-0.000020	2,985
Men	-0.000037	2,767
Women	-0.000034	2,527
Both sexes, 16 to 19 years	-0.000204	2,550
Black	-0.000125	3,139
Men	-0.000302	2,931
Women	-0.000183	2,637
Both sexes, 16 to 19 years	-0.001295	2,949
Hispanic origin	-0.000206	3,896
Not In Labor Force (use only for Total, Total Men, and White)	+0.000006	829
Agricultural Employment		
Total or White		
Men	+0.000782	3,049
Women or	+0.000858	2,825
Both sexes, 16 to 19 years		
Black	-0.000025	2,582
Hispanic origin	-0.000135	3,155
Total or Women		
Men or	+0.011857	2,895
Both sexes, 16 to 19 years		
	+0.015736	1,703
Unemployment		
Total or White		
Black	-0.000018	2,957
Hispanic origin	-0.000212	3,150
	-0.000102	3,576

[1] For not in labor force characteristics, use the Not In Labor Force parameters.

NOTE:These parameters are to be applied to basic CPS monthly labor force estimates.

For foreign-born characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born characteristics for Blacks and Hispanics.

Table 3. Parameters for Computation of Standard Errors for June 1998 Supplement Characteristics

Characteristic	Persons		Households, etc.	
	a	b	a	b
FERTILITY [1]				
Total or White	-0.000037	2,241	(X)	(X)
Black	-0.000254	2,241	(X)	(X)
Hispanic	-0.000422	4,100	(X)	(X)
Asian/Pacific Islander	-0.000444	2,241	(X)	(X)
NUMBER OF BIRTHS				
Total or White	-0.000068	4,087	(X)	(X)
Black	-0.000463	4,081	(X)	(X)
Hispanic	-0.000753	7,312	(X)	(X)
Asian/Pacific Islander	-0.000808	4,081	(X)	(X)
MARITAL STATUS, HOUSEHOLD & FAMILY CHARACTERISTICS				
Total or White	-0.000025	5,211	-0.000012	2,068
Black	-0.000300	7,486	-0.000077	1,871
Hispanic Origin	-0.000585	12,616	-0.000155	1,871
Asian/Pacific Islander	-0.000782	7,486	-0.000182	1,730
INCOME				
Total or White	-0.000012	2,454	-0.000013	2,241
Black	-0.000113	2,810	-0.000119	2,447
Hispanic Origin	-0.000220	4,736	-0.000321	4,123
Asian/Pacific Islander	-0.000294	2,810	-0.000352	2,447
EDUCATIONAL ATTAINMENT				
Total or White	-0.000011	2,369	-0.000012	2,068
Black	-0.000107	2,680	-0.000077	1,871
Hispanic Origin	-0.000084	1,811	-0.000261	3,154
Asian/Pacific Islander	-0.000226	2,164	-0.000197	1,871
NATIVITY - Born in:				
Mexico, other North America, South America	-0.000042	11,054	(X)	(X)
Europe	-0.000024	6,351	(X)	(X)
Asia, Africa, Oceania	-0.000039	10,351	(X)	(X)
Total or White	-0.000019	5,211	(X)	(X)

[1] Fertility includes number of women by number of children ever born, percent childless and women who have had a child in the last year.

Note: For foreign-born characteristics for Total and White, multiply the parameters by 1.3. No adjustments are necessary for Nativity or for foreign born characteristics for Blacks, APIs and Hispanics.

Table 4. Parameters for Computation of Standard Errors for for June 1998 Fertility Ratios

a	b	c
+0.0000014	901	1,644

Note: Multiply the parameters by 1.3 to get foreign born parameters

Table 5. State Factors to be Applied to 1998 Parameters

State	Factor	State	Factor
Alabama	1.01	Montana	0.20
Alaska	0.15	Nebraska	0.42
Arizona	0.97	Nevada	0.44
Arkansas	0.59	New Hampshire	0.38
California	1.29	New Jersey	0.82
Colorado	0.93	New Mexico	0.40
Connecticut	1.00	New York	0.89
Delaware	0.22	North Carolina	0.94
District of Columbia	0.16	North Dakota	0.16
Florida	0.97	Ohio	1.02
Georgia	1.40	Oklahoma	0.73
Hawaii	0.35	Oregon	0.86
Idaho	0.27	Pennsylvania	0.96
Illinois	1.00	Rhode Island	0.30
Indiana	1.38	South Carolina	1.01
Iowa	0.71	South Dakota	0.17
Kansas	0.65	Tennessee	1.34
Kentucky	0.92	Texas	1.21
Louisiana	0.95	Utah	0.43
Maine	0.37	Vermont	0.18
Maryland	1.38	Virginia	1.48
Massachusetts	0.81	Washington	1.47
Michigan	0.93	West Virginia	0.39
Minnesota	1.11	Wisconsin	1.23
Mississippi	0.64	Wyoming	0.12
Missouri	1.37		

Table 6. Region Factors to be Applied to 1998 Parameters

Region	Factor
Northeast	0.85
Midwest	1.03
South	1.08
West	1.09
Nonmetropolitan charactistics	1.50